VAT OF THE LEXICAL TONES IN MANDARIN CHINESE

Jiangping Kong[1], Ruifeng Zhang[1]

[1]*Center for Chinese Linguistics, Peking University*

Department of Chinese Language and Literature, Peking University

Joint Research Center for Language and Human Complexity

ABSTRACT

The purpose of this research was to investigate the association of vocal attack time (VAT) and tones in speakers of Mandarin Chinese, and to explore how tones initiated at different pitch levels affected VAT. SP and EGG signals were synchronously recorded from 72 young undergraduates or postgraduates (42 females and 30 males) while they were reading aloud a wordlist of 50 disyllabic words at their most comfortable pitch, loudness and rate. VAT measures revealed three findings. (1) Vocal attack time shows no significant difference between the common *yangping* and the *yangping* derived from *shangsheng*. This, from a physiological perspective, supports the argument that the tone sequence 3-3 in Mandarin is indeed converted into 2-3, nothing else. (2) The tones of Mandarin Chinese that start from low pitch levels (35, 21) tend to present significantly different VAT values from those that start from high pitch levels (55, 51), with mean VATs of the former being much longer than those of the latter. This embodies the nonlinear contra-variant relationship between VAT and F0 at vowel onsets. (3) There are deviations or individual differences: a small number of people do not follow this pattern.

SUBJECT KEYWORDS

Vocal attack time, Lexical tones, Phonation onset, Nonlinear contra-variant relationship

1. INTRODUCTION

1.1 VOCAL ATTACK TIME

Vocal attack time (VAT) is a concept proposed by Baken et al. (1998a, 1998b) based on the time delay between the rise of the sound pressure (SP) and the appearance of an evident electroglottographic (EGG) signal, when SP and EGG signals are recorded simultaneously. In the presence of a trans-glottal airflow, the

vocal folds oscillate with small amplitudes before they arrive at the midline of the glottis. On their arriving at the glottal midline with periodic contact achieved and stabilized, the amplitude of their oscillations grows very quickly. Therefore, the SP signal begins its growth to large magnitude well before the vocal-fold contact occurs. However, the EGG signal, as a record of vocal-fold contact area, has nearly no amplitude until the vocal-fold contact is achieved, and only after that does its amplitude grow rapidly. The EGG and SP signals are thus offset with respect to each other, and VAT is taken to be the time lag between the rise of them measured at the onset of phonation. Positive VAT values indicate that the initiation of SP signals leads that of EGG signals while negative ones signify the latter preceding the former. When the two sorts of signals rise at the same point of time, VAT equals zero. So VAT provides a potentially useful measure that varies with vocal attack characteristics. Orlikoff et al. (2009) for example, have reported negative VATs for all attempts of their subjects to produce a hard glottal attack. A computer program was developed to automatically extract VAT measures from the EGG and SP signals simultaneously recorded, and the validity of this measurement was experimentally demonstrated by Orlikoff et al. (2009). In 2012, Roark et al. proposed a figure of merit (FOM) for VAT measurement，which was actually Pearson's correlation coefficient determined from the amplitude features of SP and EGG signals (Roark et al 2012). VAT measurement has been used for nonlinguistic research by Roark et al. (2012) to acquire normative data of VAT in healthy young adults. In 2012, VAT was also measured for linguistically constrained voice onsets during the production of the six Cantonese tones (Ma et al. 2012).

1.2 LEXICAL TONES IN MANDARIN CHINESE

As a well-known tone language in Asia, Mandarin Chinese has four distinctive lexical tones: The first one, named *yinping*, is a high level tone with pitch sustained high on pitch level 5; the second, named *yangping*, is a mid-rise with pitch climbing from level 3 up to level 5; the third, *shangsheng*, is a fall-rise that dips first from level 2 to 1 and then rises to level 4; the last one, *qusheng*, is a full fall that starts from level 5 and glides all the way down to level 1; the values of these tones are consequently recorded as *yinping*(55), *yangping*(35), *shangsheng*(214), *qusheng*(51). Because these lexical tones distinguish meanings in Mandarin Chinese, the same morpheme may have different meanings when adopting different pitch contours, for example, /mi/ with a fall-rise (214) signifies 'rice' in English, but means 'honey' when its pitch contour is altered to a full fall (51).

Unlike in English where phonemic variation occurs frequently, in Mandarin Chinese there are often occurrences of Tone Sandhi. The fall-rise pitch contour of

shangsheng (214) mentioned above is only seen on syllables before pauses or in citation form. However, when two such contours are juxtaposed in speech flow, the first of them is always definitely altered into a mid-rise (35), the pitch contour that *yangping* always adopts. And furthermore, in flowing speech, the fall-rise of *shangsheng* preceding *yinping*, *yangping* or *qusheng* is nearly unexceptionally modified into a low-fall (21), with pitch dipping slightly from level 2 to 1. All these have made the original contour of Tone 3 (214) very seldom heard in connected speech. There are also occasions when Tone Sandhi is optional. The original tone of "一"，a Chinese word that means "one" in English, is 55, a high level pitch pattern, when it is in citation form or at the end of a sentence. But in flowing speech, this pattern can be modified into 35, when it precedes morphemes of *qusheng* (e.g. "一样 55+51→35+51"), or into 51, when it goes before *yinping*, *yangping* or *shangsheng* (e.g. "一般 55+55→51+55", "一直 55+35→51+35", "一起 55+214→51+214"). But not all Mandarin speakers follow suit, and there are people who still pronounce "一般" as 55+55.

1.3 PURPOSE

The aforementioned linguistic features of Mandarin tones have rendered the most frequent contours of tones in flowing speech as four types: *yinping* (55), *yangping* (35), *shangsheng* (21) and *qusheng* (51). By comparing the pitch levels from which they start, it is possible to distinguish the four types as two categories: The first and fourth tones both start with the highest pitch and go to one category, while the second and third tones that start from low levels 3 and 2 go to the other. As is well-known, pitch is a very important perceptual correlate of F0, which is associated with the rate of vocal-fold oscillations. Since the tones with a high-pitch onset have a higher rate of vocal-fold oscillation than those with a low-pitch onset during the initial stage of phonation, they may adopt different mechanisms of laryngeal adjustment, and present dissimilar characteristics of vocal attack. The purpose of the present investigation is to examine the association of VAT and tone in speakers of Mandarin Chinese, and to explore how tones initiated at different pitch levels affect vocal attack time. This is an attempt to measure VAT for linguistically constrained voice onsets.

2 METHOD

2.1 WORDLIST

Three considerations decided which disyllabic words to choose for the present study. (1) The first syllable of the word should start with a head vowel (Chao 1970), namely, no initial consonant or semivowel medial should stand at the

beginning, because the computer program designed for VAT extraction only works efficiently on syllables beginning with vowels, and a large part of a Chinese tone contour is spread on the head vowel of a syllable. (2) The three vertex vowels of Mandarin Chinese (/a/ /i/ /u/) should be chosen as the head vowels of the first syllables of words, because they occupy the utmost points on the vowel chart and represent the entire scope of tongue movement during speech. (3) Each final of the first syllable should adopt four distinctive tone patterns (*yinping*, *yangping*, *shangsheng* and *qusheng*), with each pattern being further followed by *yinping*, *yangping*, *shangsheng* and *qusheng* respectively as the second syllables of words. Since VAT measurement would only be taken for the voice onset of the first syllables, there was no requirement on the composition of the second syllables. All these resulted in a wordlist of 50 double-syllable words with /ai/ /an/ /aŋ/ /au/ /i/ /u/ chosen to be the finals of the first syllables, as is seen Table 1.

vowel of syllable1	tone combination	vowel of syllable1	tone combination	vowel of syllable1	tone combination	vowel of syllable1	tone combination	vowel of syllable1	tone combination
	哀歌 55+55		矮小 214+214		屋脊 55+214		误差 51+55		疑点 35+214
	哀求 55+35		嗳气 214+51		污垢 55+51		务实 51+35		遗憾 35+51
	哀苦 55+214	/ai/	爱惜 51+55		无边 35+55	/u/	物产 51+214		已经 214+55
	哀悼 55+51		爱情 51+35		无常 35+35		务必 51+51		以为 214+35
/ai/	挨批 35+55		碍事 51+51	/u/	无耻 35+214		一般 55+55	/i/	以免 214+214
	皑皑 35+35	/an/	安排 55+35		无故 35+51		依从 55+35		以后 214+51
	挨宰 35+214	/aŋ/	肮脏 55+55		午餐 214+55	/i/	依法 55+214		异乡 51+55
	癌症 35+51	/au/	凹版 55+214		舞池 214+35		医道 55+51		异常 51+35
	矮星 214+55		乌龟 55+55		舞蹈 214+214		疑心 35+55		意旨 51+214
	蔼然 214+35	/u/	巫婆 55+35		舞弊 214+51		遗传 35+35		议论 51+51

Table 1. 50 disyllabic words used in the present study

2.2 SUBJECTS AND INSTRUMENTATION

Recordings of the wordlist were obtained from 42 females (mean age = 24.0 years, standard deviation (SD) = 2.1) and 30 males (mean age = 22.7 years, SD = 1.9), all of whom were undergraduates or postgraduates in universities. They were able to speak standard Mandarin Chinese and use it freely for daily communication. None of them had any voice or hearing problems, and they were all in sound health at the time of testing. The process of recording was accomplished in the sound-treated booth at the Language Lab of the Chinese Department, Beijing University, where the background noise was below 25dBA. While recording, the Adobe Audition 1.5 was set at the double-channel interface with a sampling rate of 44100 Hz and a resolution of 16 bits for each channel. The electroglottograph (Model 6103) used for collecting EGG signals and the microphone and sound card

(Creative Labs Model No.sb1095) used to gain SP signals were synchronously connected to the personal computer through a sound console (Behringer XENYX502). With their lips about 10 cm away from the microphone, the subjects were asked to read aloud the disyllabic words using their most comfortable pitch, loudness and rate. Each subject read the wordlist twice, and the one reading with better quality was chosen for the present study.

Not counting the 13 bad-quality culls, the total number of speech samples eventually acquired was 3587 with 2095 from females and 1492 from males. Among the 3587 first syllables of words, 1036 were spoken with *yinping* (55), 1075 were spoken with *yangping* (35), 640 with *shangsheng* (21) and 836 with *qusheng* (51). The original tone contour 214 was absent from the database and the number of *shangsheng* tokens was smaller than others because of the Tone Sandhi described above: 214+55/35/51→21+55/35/51, 214+214→35+214. Moreover, some subjects produced the first syllable of "一般 (55+55)" as 51, while others did it as 55.

2.3 PARAMETER EXTRACTION AND DATA PREPROCESSING

VAT measures were extracted largely automatically from the EGG and SP signals using the computer program developed by Roark et al (2012), the process of which consisted of four components. The second component was to automatically identify a 600-millisecond segment of the SP and EGG signals that was centered at the approximate time of vocal onset of the first syllable of the disyllabic word. This was based on two criteria that had to be simultaneously satisfied for the band pass-filtered EGG signal: Local energy had to be greater than 15% of the maximum energy and local cycle length must have shown less than 15% variation. However, observation had shown that, for 107 speech samples from our database, the 600-millisecond segments thus identified were not centered at the voice onset of the first syllables, but somewhere else, for example at the onset of the second syllables, suggesting that inadequate EGG signal quality of these samples failed the two criteria above. VAT measures of these samples were marked "WS" (wrongly segmented) in the comments column of excel sheets.

From the 3587 VAT values obtained, the 107 "WS" measures were first taken away, and the remaining 3480 were divided into two groups: 2025 measures for females, and 1455 measures for males. Each group was then processed separately in the same way: Because of the large number of outliers among VAT values, measures that were beyond ±2 standard deviations from the mean VAT were cut out. Eventually, another 100 measures were deleted from the database, and among the 3380 that remained (skewness =-0.32, kurtosis =3.397), VAT ranged from -40.4 ms to 37.1 ms, the average and standard deviation of it being -0.32ms and 7.16 ms

respectively. SPSS 13.0 (SPSS. Inc. USA) was used for all the analyses below.

3 RESULTS

Among the preprocessed database, there are 1000 speech samples, whose first syllables carry mid-rise pitch contours of *yangping* (35), and of these contours, 195 are derived from *shangsheng* (214) by this pattern of Tone Sandhi: 214+214→35+214. A one-sample *t* test showed that the VAT values of these 195 derived *yangping* contours are not significantly different from those of the common *yangping* contours ($t_{(194)}$ =1.486, p = 0.139 > 0.05). Similarly, there were 43 subjects, who pronounced the first syllable of "一般 (55+55)" as 51, a full fall, and according to another one-sample *t* test, the VAT values of these 43 derived *qusheng* pitch contours (51) are not significantly different from those of the common *qusheng* contours either ($t_{(42)}$ = 0.822, p = 0.416 > 0.05). It is therefore reasonable to regard these derived pitch patterns of 35 and 51 as belonging to common *yangping* and *qusheng* categories respectively.

Since the final data corpus is composed of 1961 VAT measures for females, 1419 for males, 992 for *yingping* (55), 1000 for *yangping* (35) (including 195 derived ones), 599 for *shangsheng* (21) and 789 for *qusheng* (51) (including 43 derived ones), a two-way analysis of variance with mixed measures on two factors was done with speaker gender (female vs. male) as the between-subject factor and tone (the 4 tones) as the within-subject factor. Results revealed a significant main effect of tone ($F_{(2.444)}$ = 33.59, p < 0.05, R^2 = 0.324), a non-significant main effect of gender ($F_{(1,70)}$ = 0.179, p = 0.673 > 0.05, R^2 = 0.003), and a non-significant tone by gender interaction effect ($F_{(2.444)}$ = 0.262, p = 0.813 > 0.05, R^2 = 0.004). Means and SDs of VAT calculated among the 72 subjects are listed in Table 2 and plotted in Figure 1.

	Sex	Mean	Std. Deviation	N
yinping(55)	female	-1.602	5.406	42
	male	-1.710	4.235	30
	Total	-1.647	4.920	72
yangping(35)	female	1.573	3.878	42
	male	2.176	3.675	30
	Total	1.824	3.781	72
shangsheng(21)	female	.903	4.826	42
	male	1.612	4.812	30
	Total	1.198	4.799	72
qusheng(51)	female	-2.462	6.595	42
	male	-2.017	3.576	30
	Total	-2.276	5.512	72

Table 2. Means and SDs of VAT of different tones and genders among all the 72 subjects

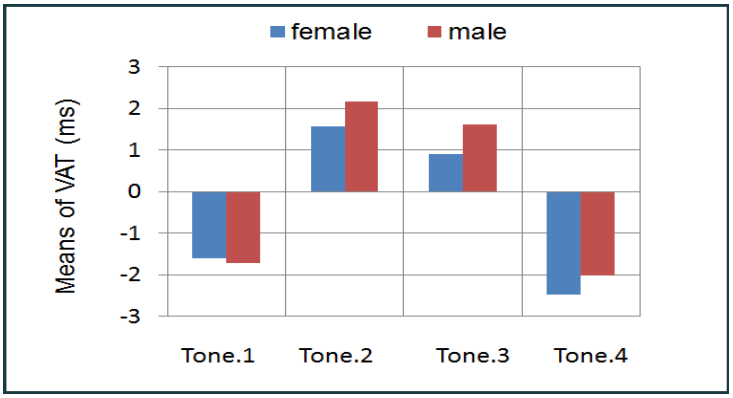

Figure 1. Means of VAT as a function of tone and gender (72 subjects)

Figure 1 indicates that the average VATs for three of the four tones are all smaller for females than for males, but those for tone1 are not, suggesting the necessity for analyses of simple effect. Thus a paired-samples *t* test has further shown that, for both males and females, VATs between *yinping* (55) and *yangping* (35), *shangsheng* (21) and *qusheng* (51) are significantly different from each other (For males: 55 vs.35: $t = -5.421$, $p = 0.00 < 0.05$; 21 vs. 51: $t = 4.479$, $p = 0.00 < 0.05$. For females: 55 vs.35: $t = -4.526$, $p = 0.00 < 0.05$; 21 vs. 51: $t = 4.291$, $p = 0.00 < 0.05$), while those between *yangping* (35) and *shangsheng* (21), *yinping* (55) and *qusheng* (51) are not (For males: 35 vs. 21: $t = 1.007$, $p = 0.322 > 0.05$; 55 vs. 51: $t = 0.702$, $p = 0.488 > 0.05$; For females: 35 vs. 21: $t = 1.222$, $p = 0.229 > 0.05$; 55 vs. 51: $t = 1.34$, $p = 0.188 > 0.05$). Namely, the means of VAT are much longer for *yangping* and *shangsheng* than for *yingping* and *qusheng* in both genders. This is why a cluster analysis has put *yinping* and *qusheng* into one category, but *yangping* and *shangsheng* into another depending on the VAT measures of their voice onsets (see Figure 2).

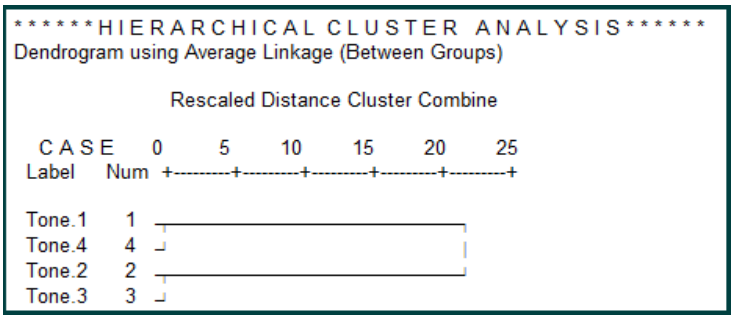

Figure 2. Result of a hierarchical cluster analysis

However, a close inspection on the mean VATs of the four tones (55, 35, 21, and 51) of each subject can also divide the 72 subjects (including males and females)

into two groups: 46 of them (63.89%) have mean VATs of both *yangping* and *shangsheng* longer than those of *yinping* and *qusheng* (see Figure 3), but 26 of them (36.11%) display various patterns other than this (see Figure 4).

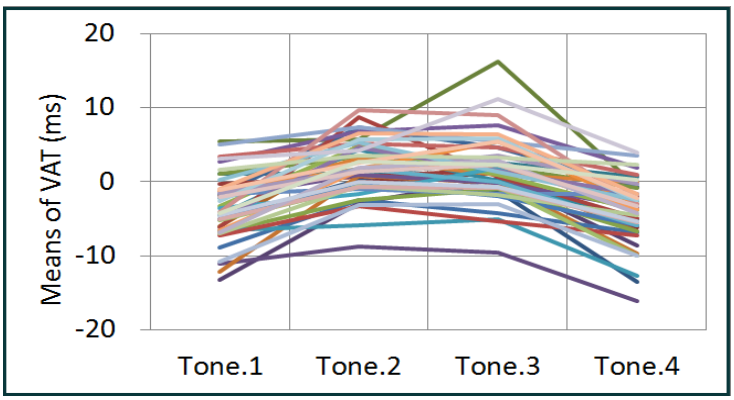

Figure 3. 46 of the 72 subjects displayed mean VATs of four tones in the same pattern. Each line indicates the average VATs of one person

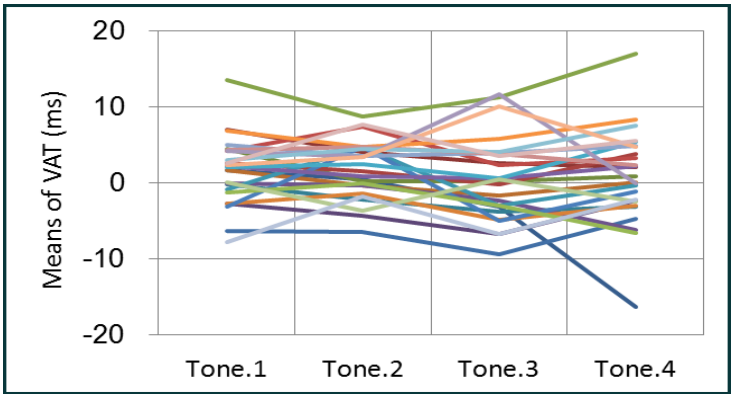

Figure 4. 26 of the 72 subjects displayed mean VATs of four tones in miscellaneous patterns. Each line indicates the average VATs of one person

A two-way analysis of variance with mixed measures on two factors test done on the 46 still presents a significant main effect of tone ($F_{(3)} = 87.644$, $P < 0.05$, $R^2 = 0.666$), a non-significant main effect of gender ($F_{(1,44)} = 2.163$, $p = 0.149 > 0.05$, $R^2 = 0.047$) and a non-significant tone by gender interaction effect ($F_{(3)} = 0.632$, $P = 0.595 > 0.05$, $R^2 = 0.014$). In Table 3 and Figure 5 are indicated the means and SDs of VAT of the 46 subjects as a function of tone and gender. The exception in Figure 1 no longer shows up in Figure 5, suggesting that individual differences may slightly affect the result of the whole population.

	Sex	Mean	Std. Deviation	N
yinping(55)	female	-3.847	4.641	26
	male	-2.992	3.746	20
	Total	-3.475	4.251	46
yangping(35)	female	1.199	4.193	26
	male	2.751	3.147	20
	Total	1.874	3.815	46
shangsheng(21)	female	.807	4.781	26
	male	2.778	3.630	20
	Total	1.664	4.386	46
qusheng(51)	female	-4.839	4.621	26
	male	-2.811	3.713	20
	Total	-3.957	4.326	46

Table 3. Means and SDs of VAT of different tones and genders among the 46 subjects

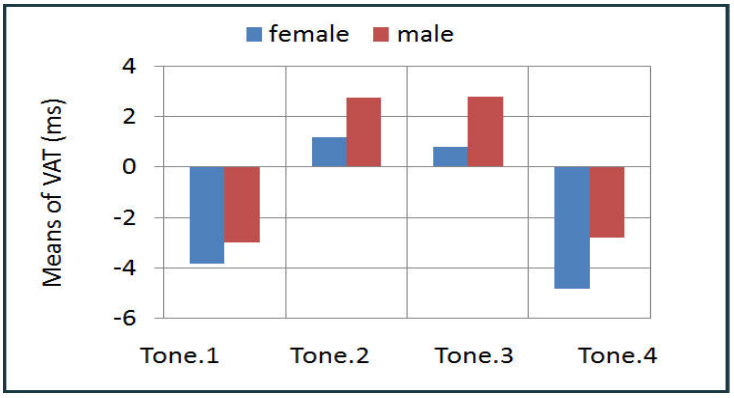

Figure 5. Means of VAT as a function of tone and gender (46 subjects).

4. DISCUSSION

4.1 TONE SANDHI

One debate concerning tone sandhi in Mandarin Chinese had been whether the tone sequence 3-3 is homophonous with the sequence 2-3. The issue was eventually settled by Wang et al (1967, 2006) with a perception experiment. The 130 pairs of test items used in his research were thus designed: The two members of each pair shared the same phonological features except that of pitch contour; In other words, one member carried the tone sequence 2–3 while the other carried 3–3. These items were recorded in random order and then presented randomly for native speakers of Mandarin to listen to and identify whether they were sequences 3-3 or 2-3. None of the listeners were able to present a rate of accuracy over 55%, and even the person from whom the test items had been recorded couldn't correctly identify over 60% of them, suggesting that the *yangping* pitch contour (35) derived from *shangsheng* (214) was perceptually no different from that of the common *yangping*

(35). The result of the first one-sample *t* test done in the present study supports this argument from the perspective of physiology: VAT values of the second tone derived from the third tone are not significantly different from those of the common second tone, suggesting that all the *yangping* contours, whether original or derived, share similar laryngeal adjustments at their voice onsets, and therefore display approximate features of vocal attack. In a word, both perception and physiology point to one conclusion: The pitch contour 214 before another 214 is indeed phonemically the same as *yangping*. The result of the second one-sample *t* test leads to a similar judgment： The onsets of phonation of all the *qusheng* contours, whether the original 51 or the ones derived from 55, are physiologically alike, and the two kinds should be perceptually no different.

4.2 VAT AND LEXICAL TONES

The second finding from the analyses above can be summarized as follows: In a large group of Mandarin speakers, the two lexical tones with high-pitch onsets, *yinping* (55) and *qusheng* (51), display smaller VAT values, but the other two with low-pitch onsets, *yangping* (35) and *shangsheng* (21), present much larger ones (see tables 1 and 2 and figures 1, 2, 3 and 5); In other words, a higher rate of vocal-fold oscillation tends to be associated with a shorter VAT value, and vice versa. This negative VAT-F0 correlation at the linguistically constrained voice onset is also seen in the three level tones of Cantonese (Ma et al., 2012): In females, mean VATs of high, mid and low level tones are respectively 0.72 ms, 1.70 ms, 1.78 ms; In males, mean VATs of high, mid, low level tones, although longer than those in females, are also thus lined up, 3.99ms, 4.64ms, 4.69ms. However, Tables 2 and 3 also indicate counterexamples: For both males and females, mean VATs of *shangsheng* (21) should always be larger than those of *yangping* (35), because the former has a lower initial pitch than the latter; but this is actually not the case. These conform to the finding of the VAT study on 5 linguistically unconstrained pitch levels in Mandarin (Zhang et al). In a large group of people, as pitch levels shift from one to five, there is a linear increase of pitch, but a nonlinear decrease of VAT: from Levels Two to Five, each mean value of VAT is not always larger than the one that follows; But, the average VAT at Level One is always the largest among the five pitch levels, and is much larger than that of all the others. Therefore, for both linguistically constrained and unconstrained vocal onsets, VAT and pitch tend to present a nonlinear contra-variant relationship in most mandarin speakers.

4.3 INDIVIDUAL DIFFERENCS

46 of the 72 subjects produced the low-pitch onsets of the second and third tones (35, 21) with longer VAT means than they did the high-pitch onsets of the first and fourth

tones (55, 51), while 26 of them showed inconsistent patterns of VAT means in pronouncing the four. This seems to support the findings by Zhang et al. that as pitch means increase linearly from Levels One to Five, mean VATs decrease nonlinearly in a large group of people but increase nonlinearly in a small group of them, and that different people incline to use different strategies in increasing pitch height. However, among the 26 subjects observed in the present study, mean VATs of four tones are ordered as *yangping* (35) 1.736 ms > *yinping* (55) 1.586ms > *qusheng* (51) 0.697 ms > *shangsheng* (21) 0.375 ms, and a positive VAT-F0 correlation is not seen at the phonation onsets of the four tones. What caused the individual differences needs to be further investigated.

5 CONCLUSION

Firstly, vocal attack time, as a measure of phonatory function of the vocal folds, shows no significant difference between the common *yangping* and the *yangping* derived from *shangsheng*, and between the common *qusheng* and the *qusheng* derived from *yinping*. This is physiologically in support of the argument that the tone sequence 3-3 in Mandarin is indeed converted into 2-3, nothing else. Secondly, the tones of Mandarin Chinese that start from low pitch levels (35, 21) tend to present significantly different VAT values from those that start from high pitch levels (55, 51), with mean VATs of the former being much longer than those of the latter. This is with the nonlinear contra-variant relationship between VAT and F0 at the vowel onsets. Thirdly, there are deviations or individual differences: a small number of people do not follow this pattern.

NOTES

1.This research was funded by the National Social Sciences Foundation of China. Grant No:10&ZD125.

REFERENCES

- BAKEN R.J. and Orlikoff R.F. 1998a. Vocal fold adduction time estimated from glottographic signals. Presented at: the 25th Mid-Winter Meeting of the Association for Research in Otolaryngology; February 1998; St. Petersburg, FL.
- _____.1998b. Estimating vocal fold adduction time from EGG and acoustic records. In: Schutte HK, Dejonckere P, Leezenberg H, Mondelaers B, HF, eds. Programme and Abstract Book: 24th IALP Congress, Amsterdam; 1998:15.
- Ma E. Baken R.J. and Roark R.M. 2012. Effect of Tones on Vocal Attack Time in

Cantonese Speakers. Journal of Voice, Vol. 26, issue. 5, pp. 670.e1-670.e6.

- ROARK R.M. Watson B.C. and Baken R.J. 2012. A Figure of Merit for Vocal Attack Time Measurement. Journal of Voice. Vol.26, issue 1, pp.8-11.
- _____. Watson B.C. Baken R.J. Brown D.J. et al. 2012. Measures of Vocal Attack Time for Healthy Young Adults. Journal of Voice. Vol.26, issue 1, pp.12-17.
- ORLIKOFF, R.F. Deliyski D.D. Baken R.J. and Watson B.C. 2009. Validation of a Glottographic Measure of Vocal Attack. Journal of Voice. Vol 23, issue 2, pp.164-168.
- WANG W. S-Y. Li, K.P. 1967. Tone 3 in Pekinese. Journal of Speech Hearing Research. 10 (3), pp. 629-36.
- ____. Peng Gang. 2006. Language, Speech and Technology. Shanghai Educational Publishing House. pp. 120-121.
- CHAO Yuen Ren. 1970. A Grammar of Spoken Chinese. University of California Press, Berkeley, Los Angeles, London. pp. 18-25.
- ZHANG Ruifeng, Baken R.J, Kong Jiangping. Vocal Attack Time of Different Pitch Levels and Vowels in Mandarin. Journal of Voice (in press).

普通話四聲的 VAT 研究

孔江平　張銳鋒
北京大學中國語言文學系
中國語言學研究中心
語言與人類複雜系統聯合研究中心

提要

漢語作為典型的聲調語言，有著四個能區分意義的詞調：陰平(55)，陽平(35)，上聲（214），去聲（51）。但由於連讀變調，在語流中經常出現的四聲調值為：陰平（55），陽平（35），上聲（21）和去聲（51）。本文旨在研究漢語普通話者的 VAT 與聲調的關係，並探索起始於不同音高層級的聲調對 VAT 的影響情況。我們從 72 位發音人(男 30 人和女 42 人)同時錄製了 50 個雙音節詞的 SP 和 EGG 信號，所有發音人均為二十多歲的在校大學生或研究生。單獨樣本的 T 檢驗表明：陽平及由上聲變來的陽平，去聲及由陰平變來的去聲在 VAT 值上不存在顯著性差異。這就從生理上支持了兩上聲相連，前上變陽平的觀點。二因素重複測量的方差分析表明：起始於低音高層級的聲調的 VAT 值與起始於高音高層級的聲調的 VAT 值存在顯著性差異，前者的 VAT 值明顯大於後者。但還存在個體差異，72 位發音人中有 26 位不符合這一模式。

關鍵字
聲門碰撞時間　字調　發聲起始　非線性反變關係